More computing resources

- Running R in batch mode
- ssh and remotely accessing a computer
- Amazon Web Services (AWS)
- Hadoop

running batch scripts in R

```
R CMD BATCH --vanilla --slave test.R  &
```

`--vanilla` - no saved workspaces or user defined environment
`--slave` - keep R quiet
`&` - run in the background

you have to save output, plots, and data in your script

An alternative:

```
Rscript --vanilla  test.R & > test.Rout &
```

On Windows (at the command line)

```
"C:\Program Files\R\R-2.13.1\bin\R.exe" CMD BATCH
   --vanilla --slave "c:\my projects\my_script.R"
```

Not a huge improvement over just opening another R session and sourcing your code there

UNLESS you are running on a remote computer, and only have a command line interface,

but if you log out, your process will be killed…

ssh and remote computers

ssh

ssh is a protocol for communicating securely with a remote computer.

Mac & Linux should have ssh by default (on the command line)

`ssh wickhamc@app.science.oregonstate.edu`

Windows

- get PuTTY
- start new connection with host: app.science.oregonstate.edu and defaults

access `app.science.oregonstate.edu` with your science username and password and you will have access to your Z drive. Has R and git.

nohup & screen

The remote computer is probably linux based (? they always have been for me). `nohup` lets you start a process and logout and the process will keep running.

```
nohup R CMD BATCH --vanilla --slave test.R  &
```

Check out `screen` too if you are interested.

AWS

Amazon web services

Amazon web services is a suite of services that let you buy computing resources.

In particular EC2 - allows you to start up your own virtual computer. It costs $ - but there is a free tier.

Amazon Machine Images (AMI) are combinations of operating systems and installed software that allow you to start a machine in a desired state.

Check out: http://www.louisaslett.com/RStudio_AMI/ for an AMI with R, RStudio and git

Getting AWS set up
Getting started with EC2

- start EC2 instance with RStudio AMI
- connect via browser
- new project from git

Demo Rstudio server

Careful, if you terminate your instance, you lose all your data. Copy it off, or commit and push to git before terminating the instance.

For more control, use ssh instead:

- gives you command line access to machine so you can set up users, passwords, install other software etc.
- Run things in batch mode and logout (if you use nohup)

Hadoop

Hadoop

HDFS + MapReduce + glue that makes it all easy to deploy and fault tolerant

HDFS - a distributed file system
MapReduce - a split-apply-combine paradigm

Big Idea: spread data and computation across multiple computers. Keep computation local so data isn't being moved around.

R interfaces
`RHadoop`
`RHipe`